# EXPERIMENT 9

**A.1 Aim:**

To study IDS and the various performance measures of IDS (Accuracy, Precision, F1. Score)

**A.2 Prerequisite:**

Fundamentals of IDS

Tools: Weka, NSL KDD IDS dataset

**A.3 Outcome:**

**After successful completion of this experiment students will be able to**

1. Understand False Positives, True Positives, False Negatives, True Negatives.

**Theory:**

True positive (TP): Its value represents the number of malwares that have been correctly classified as malwares.

False negative (FN): Its value represents the number of malwares that have been misclassified as normal programs.

False positive (FP): Its value represents the number of normal applications that have been misclassified as malwares.

True negative (TN): Its value represents the number of normal applications that have been correctly classified as normal.

**Task 1:**

Download the NSL KDD dataset and use weka to study the features. Prepare a descriptive statistics using weka.

**Task 2.** Compare the performance measures using Random Forest and Naïve Bayes Algorithm

Note the values of TP, FP, FN, TN, Accuracy, Precision and F.1 score in both the cases.

# PART B

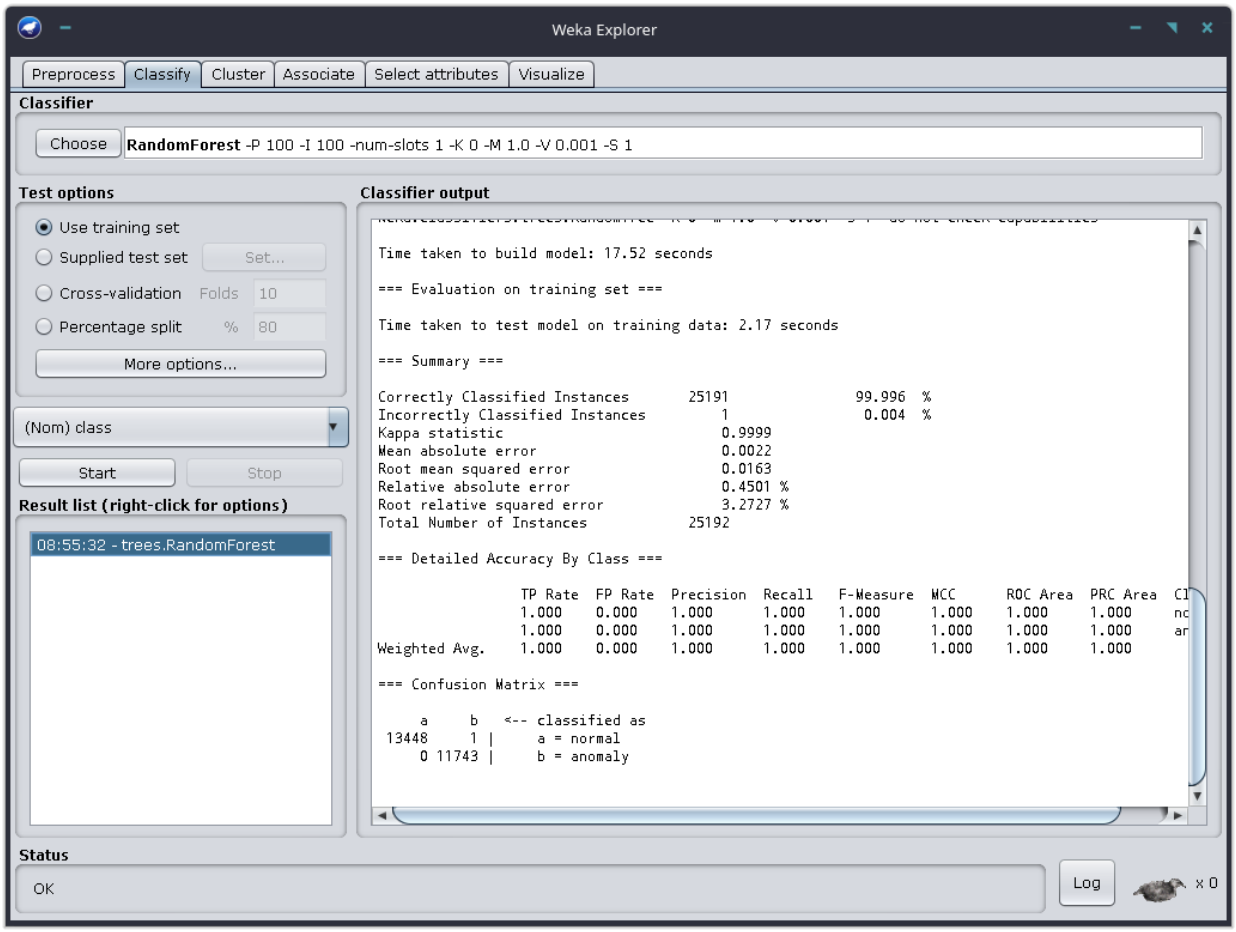
(PART B : TO BE COMPLETED BY STUDENTS)

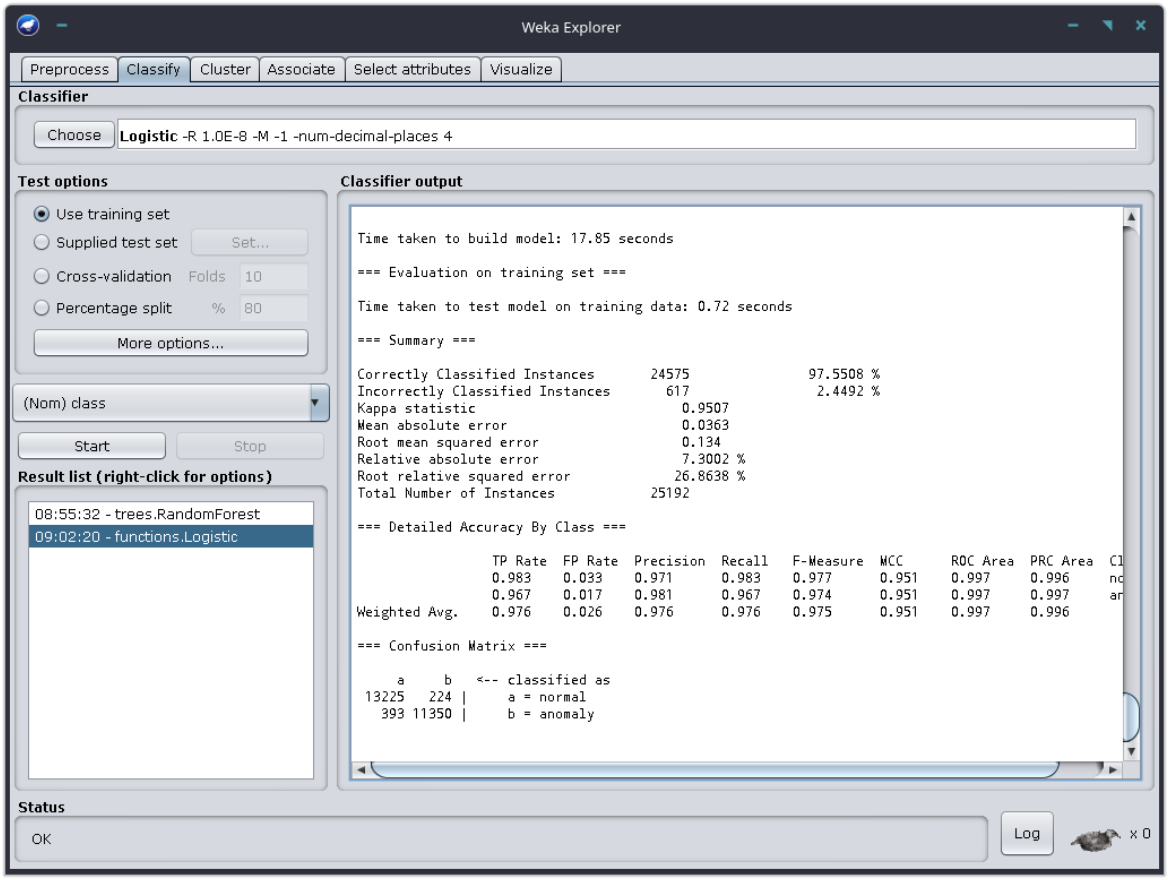
***(Students must submit the soft copy as per following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case the there is no Black board access available)***

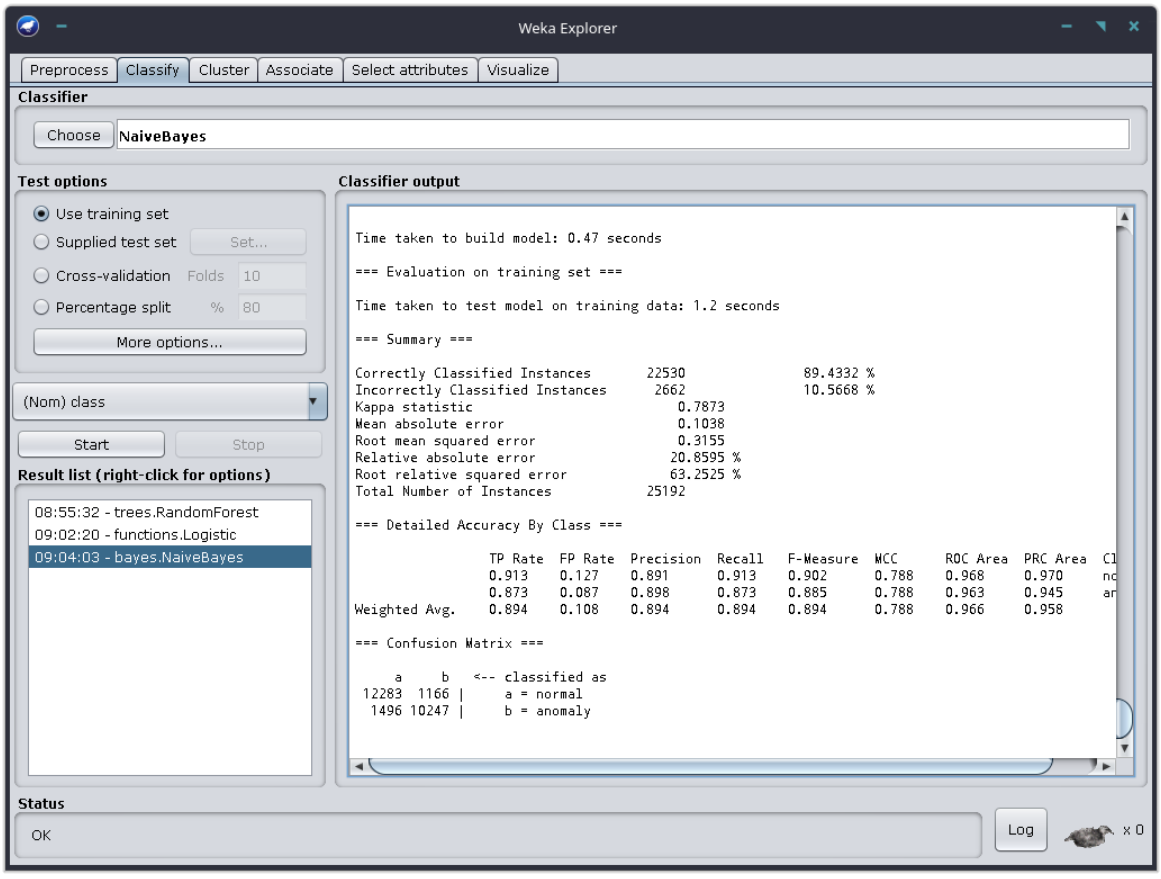
| | |
|---|---|
| Roll. No. N233 | Name: Hrushit Jain |
| Class: MBA Tech CS | Batch: G |
| Date of Experiment: 11/9/2020 | Date of Submission: 11/9/2020 |
| Grade: | |

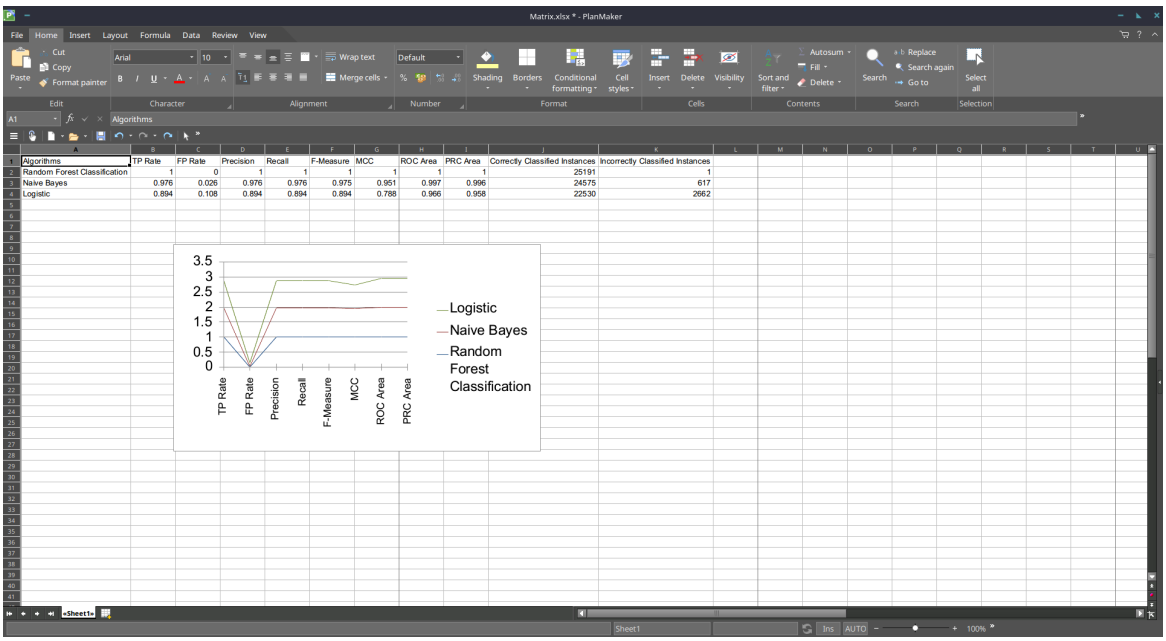
## B1. Output:

Take the screenshots for the Task1 and Task2




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Logistic -R 1.0E-8 -M -1 -num-decimal-places 4
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
80
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
08:55:32 - trees.RandomForest
09:02:20 - functions.Logistic
Classifier output
Time taken to build model: 17.85 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0.72 seconds
=== Summary ===
Correctly Classified Instances 24575 97.5508 %
Incorrectly Classified Instances 617 2.4492 %
Kappa statistic 0.9507
Mean absolute error 0.0363
Root mean squared error 0.134
Relative absolute error 7.3002 %
Root relative squared error 26.8638 %
Total Number of Instances 25192
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area
0.983 0.033 0.971 0.983 0.977 0.951 0.997 0.996
0.967 0.017 0.981 0.967 0.974 0.951 0.997 0.997
Weighted Avg. 0.976 0.026 0.976 0.976 0.975 0.951 0.997 0.996
=== Confusion Matrix ===
a b <-- classified as
13225 224 | a = normal
393 11350 | b = anomaly
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
80
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
08:55:32 - trees.RandomForest
09:02:20 - functions.Logistic
09:04:03 - bayes.NaiveBayes
Classifier output
Time taken to build model: 0.47 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 1.2 seconds
=== Summary ===
Correctly Classified Instances 22530 89.4332 %
Incorrectly Classified Instances 2662 10.5668 %
Kappa statistic 0.7873
Mean absolute error 0.1038
Root mean squared error 0.3155
Relative absolute error 20.8595 %
Root relative squared error 63.2525 %
Total Number of Instances 25192
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area
0.913 0.127 0.891 0.913 0.902 0.788 0.968 0.970
0.873 0.087 0.898 0.873 0.885 0.788 0.963 0.945
Weighted Avg. 0.894 0.108 0.894 0.894 0.894 0.788 0.966 0.958
=== Confusion Matrix ===
a b <-- classified as
12283 1166 | a = normal
1496 10247 | b = anomaly
Status
OK
Log
x 0




| Algorithms | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Correctly Classified Instances | Incorrectly Classified Instances |
|---|---|---|---|---|---|---|---|---|---|---|
| Random Forest Classification | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 25191 | 1 |
| Naive Bayes | 0.976 | 0.026 | 0.976 | 0.976 | 0.975 | 0.951 | 0.997 | 0.996 | 24575 | 617 |
| Logistic | 0.894 | 0.108 | 0.894 | 0.894 | 0.894 | 0.788 | 0.966 | 0.958 | 22530 | 2662 |

## B.2 Observations and learning:

***(Students are expected to comment on the output obtained with clear observations and learning for each task/ sub part assigned)***

**We learn about the various ML algorithms and their application in security and Intrusion Detection Systems.**

## B3: Questions of Curiosity

Give few examples of commercially used IDS. How do they differ from each other.

Bro IDS, Snort, Ethereal, Prelude, Multi Router Traffic Grapher and Tamandua network based **IDS**, and then give a collection of existing available **commercial** IDSs products.

## B.4 Conclusion:

***(Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)***

**We learn about all the IDS and latest methods/technologies applied to the field of security.**